Assessing Physical Rehabilitation Exercises using Deep Learning

Mrigank Shah, Sanjana Krishnam and Tarun Anand

***Abstract*— The scope for assessing physical rehabilitation by evaluating the performance of patients while completing the required rehabilitation exercises is immense. We can do this by capturing and processing the movement data from recuperating patients. Rehabilitation assessment plays an important role in reducing healthcare costs and furthering patients wellness. Despite its major role, current approaches for computer aided performance evaluation is extremely limited in terms of versatility and robustness. We aim to extend the work done by [1] . The deep learning framework suggested by [1] which included metrics for quantifying movement performance, scoring functions, autoencoder neural networks and deep neural network models for regressing quality score have been enhanced and improved. The main components of the framework are metrics for quantifying movement performance, scoring functions for mapping the performance metrics into numerical scores of movement quality, and deep neural network models for regressing quality scores of input movements via supervised learning.**

***Index Terms*—Movement Modeling; Deep Learning Convolutional Neural Networks;**

I. Introduction

Physical therapy and rehabilitation programs form an important part of a patient's recovery process. It is a necessary treatment for a wide array of debilitating conditions. However, it is prohibitively expensive and economically unjustified to offer patients access to a clinician for every single rehabilitation session [2]. Several factors have been identified that contribute to the low compliance rates, but the most important factor is the absence of continuous feedback and timely oversight of patient's exercises in a home environment by a healthcare professional [3].Despite the development of a variety of new tools and devices in support of physical rehabilitation, such as robotic assistive systems [4], virtual reality and gaming interfaces [5],and Kinect-based assistants [6], [7], there is still a lack of versatile and robust systems for automatic monitoring and assessment of patient performance.

The first mention of a deep learning based approach to tackle this issue was suggested in [1]. Our approach has improved upon the same and yields a higher accuracy than [1]. This framework includes formulation of metrics for quantifying movement performance, scoring functions for mapping the performance metrics into numerical scores of movement quality, and deep learning-based end-to-end models for encoding the relationship between movement data and quality scores.

The studied performance metrics are classified into model-less and model-based groups of metrics [8]. The model-less metrics employ distance functions, such as Euclidean distance and dynamic time warping (DTW) [9] deviation between data sequences. The model-based metrics apply probabilistic approaches for modeling the movement data, and consequently, employ the log-likelihood for performance evaluation [10]. Next, we investigate the effectiveness of deep autoencoder networks for dimensionality reduction of captured data. Further, we propose scoring functions for scaling the values of the studied performance metrics into the [0, 1] range. The resulting movement quality scores are employed as the ground truth for training the proposed neural networks (NNs) for rehabilitation applications.

The proposed framework compares the performance of convolutional neural networks (CNNs), recurrent neural networks (RNNs), and hybrid neural networks (HNNs). The framework is validated on the University of Idaho Physical Rehabilitation Movement Dataset (UI-PRMD)[11].

The paper is organized as follows. The next section describes the Literature Review which is a compilation of similar works. Section III describes the various performance metrics that have been tried and implemented. Section IV describes the framework implemented. Section V discusses the results obtained and Section VI is our conclusion

II. Literature Review

There have been several studies on exercise evaluation using machine learning models to classify exercises as correct or incorrect. Models used for this include k nearest neighbours [12] and multilayer perceptron NNs [13]. However the outputs are binary values and do not provide the capacity to detect varying levels of movement quality.

There were studies that employed distance functions for deriving movement quality scores. [14] makes use of the Mahalanobis distance to quantify the exercise performed. Dynamic Time Warping was used to calculate the distance between patient performed repetitions and a healthy subject in [15]. These approaches have shortcomings because there is no model that is generated and cannot be utilized for other exercises.

III. Performance Metrics

Performance metrics evaluate the level of correctness of each repetition with respect to the set of reference movements. Considering the fact that the existing datasets of movements contain data collected by multiple subjects, the performance metrics can be based on the between subject or within subject case.

A. Euclidean Distance

Euclidean distance: the Euclidean distance between two movement data $Y_{r,s}$ and $X_{r',s'}$ has been used and the Euclidean distance is defined by the common Euclidean Distance formula.

B. DTW Distance

DTW Distance: Dynamic time warping (DTW) algorithm is an algorithm for aligning time-series data via nonlinear warping of the temporal order of the data points in order to reduce a distance function between the time-series. The most commonly used distance function in DTW is the Euclidean distance. The implementation of DTW is described in [9].

C. Within and Between subjects metrics

Within Subjects: Within subjects metrics calculate the deviation between a repetition of an exercise $Y_{r,s}$ and a set of repetitions of the exercise performed by the same subject.

Between Subjects: Between subjects metrics calculate the deviation between a repetition of an exercise $Y_{r,s}$ and a set of repetitions of the exercise performed by a different subject.

D. Separation Degree

Separation Degree: The separation degree indicates greater ability of the used metric to differentiate between correct and incorrect repetitions of an exercise. For comparison of the scaled values of the performance metrics we propose the concept of separation degree. Specifically, for any positive real numbers x, y their separation degree is defined as S_D(x,y) = $\frac{x-y}{x+y}$ ϵ [-1,1]. The separation degree between two positive sequences x = $(x_1, x_2, ...x_L)$ and y = $(y_1, y_2, ...y_L)$ is defined by

$$S_D(x,y) = \frac{1}{mn} \sum_{i=1}^{m} \sum_{i=1}^{n} S_D(x_i, y_j) \tag{1}$$

Values of the separation degree close to -1 or 1 indicate good separation between the two sequences. Conversely, for values of the separation degree close to 0, the sequences dont separate well and they are almost mixed together

IV. Model Implemented

The framework that we have implemented comprises of dimensionality reduction, performance metrics, scoring functions and NN models. The measured joint coordinates data by the sensory system are processed via dimensionality reduction, performance metric, and scoring function to obtain movement quality scores that are used for training a NN model. The trained NN model is afterward used to automatically generate movement quality scores for input movement data acquired by the sensory system. Figure 1 provides an overview of the working of our model.

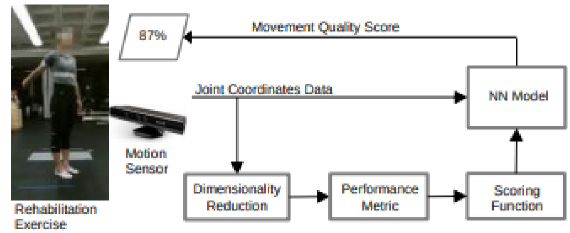

Fig. 1. Framework implemented

A. Dimensionality Reduction

The outputs of the sensory system for Outputs from the sensory systems for capturing human joint displacements are high-dimensional data, typically ranging between 40 and 120 dimensions. Dimensionality reduction of recorded data is often considered an essential step in processing human movements, in order to remove redundant, or highly correlated dimensions. The goal is to project the data into a lower dimension representation. Principal Component Analysis(PCA) is one of the techniques utilized for dimensionality reduction, where a matrix containing the leading M eigenvectors corresponding to the largest eigenvalues of the covariance matrix V is used for projecting the data into a lower-dimensional space.

One of the shortcomings of PCA is that it employs a linear mapping of high-dimensional data into a lower-dimensional representation. KPCA is a variant of PCA that uses a kernel trick to extend conventional principal component analysis to a high dimensional feature space.

We implemented three types of KPCA's:

1) Linear
2) Polynomial
3) Rbf

We implemented autoencoder neural networks [16]. Autoencoder NNs is a nonlinear technique for dimensionality reduction, which allows extracting richer data representations for dimensionality reduction in comparison to the linear techniques (such as PCA). Furthermore, deep autoencoder NNs created by stacking multiple consecutive layers of hidden neurons, can additionally increase the representational capacity of the network.

Autoencoders are an unsupervised form of NNs designed to learn an alternative representation of input data, through a process of data compression and reconstruction. The data processing involves an encoding step of compressing input data through one or multiple hidden layers, followed by a decoding step of reconstructing the output from the encoded representation through one or multiple hidden layers.

B. Performance Metric

The performance metric that we implemented was DTW. This choice was based on the accuracy obtained when compared to other performance metrics.

C. Scoring function

Scoring functions map the value of the performance metrics into a movement quality score in the range between 0 and 1. The resulting movement quality scores play a dual role in the proposed framework. First, in a real-world exercise assessment setting, the quality scores allow for intuitive understanding of the calculated values of the used performance metric. For instance, a movement quality score of 88% presented to a patient is easy to understand, and it can also enable the patient to self-monitor his/her progress toward functional recovery based on received quality scores over a period of time. Second, the quality scores are used here for supervised training of the studied NN models.

For a sequence of metric values of the reference movements x = $(x_1, x_2, ...x_L\ )$ and a sequence y = $(y_1, y_2, ...y_L)$ related to the patient movements, we propose the following scoring functions:

$$\bar{x_k} = (1 + e^{\frac{x_k}{\mu+3\delta} - \alpha})^{-1} \tag{2}$$

$$\bar{y_k} = (1 + e^{\frac{x_k}{\mu+3\delta} - \alpha} + \frac{y_k - x_k}{\alpha(\mu + 3\delta)})^{-1} \tag{3}$$

The value of α is a data specific parameter. The proposed scoring function is monotonicaly decreased.

D. Neural Networks

Three different deep NN architectures are developed, implemented and evaluated in this work. These include CNNs, RNNs, and a hybrid model combining CNN and LSTM. Various combinations of layers, numbers of layers, computational units per layer, size of convolutional filters, batch size, and other related hyperparameters were tried until we arrived at an optimal combination. For all models, mean-squared-error was selected as a cost function, and Adam optimizer was employed. A batch size of 5 was applied, with early stopping regularization. Inputs are 117-dimensional sequences of joint displacements corresponding to single repetitions of an exercise. The output layer has linear activations, and outputs a numerical movement quality score for an input repetition.

The adopted CNNs contain five convolutional layers, two fully connected hidden layers, and an output layer. They utilize strided one-dimensional convolutional filters, leaky ReLU activations, and dropout of 0.2.

The RNN models with recurrent architecture consist of two bidirectional layers of LSTM units, one intermediate full connected layer, and an output layer. The recurrent layers use a recurrent dropout of 0.5, and are as well followed by a dropout layer with the rate of 0.25.

The hybrid model contains 6 convolutional layers followed by a bidirectional layer of LSTM units, two fully connected hidden layers, and an output layer. The recurrent layers use a recurrent dropout of 0.5, followed by a dropout layer with the rate of 0.25. The other layers use ReLU activations and a dropout of 0.2.

V. Experimental Results

In this section we evaluate the performance of our system.

A. Separation degree performance

After applying the DTW algorithm to the between and within subjects results of the various dimensionality reduction methods, the separation degree was calculated for each method. The results are as in TABLE I.

TABLE I: SEPARATION DEGREE RESULTS

Model	Performance Metric	Separation Degree
Autoencoder	Between Subjects	-0.448
Autoencoder	Within Subjects	-0.747
PCA	Between Subjects	-0.302
PCA	Within Subjects	-0.703
Linear-KPCA	Between Subjects	-0.302
Linear-KPCA	Within Subjects	-0.703
Poly-KPCA	Between Subjects	-0.420
Poly-KPCA	Within Subjects	-0.741
Rbf-KPCA	Between Subjects	-0.680
Rbf-KPCA	Within Subjects	-0.815

The within-subject case provides improved separation because the repetitions performed by the same subject are characterized. Since the rbf-KPCA method has the highest separation degree value, we choose the rbf-KPCA method to continue. Only the case of between-subject is considered, because for the within-subject cases the number of repetitions per subject is too low for NNs training and the between-subjects metric enhances variability.

The DTW vs sequence number between subjects for the rbf KPCA is plotted in Figure 2.

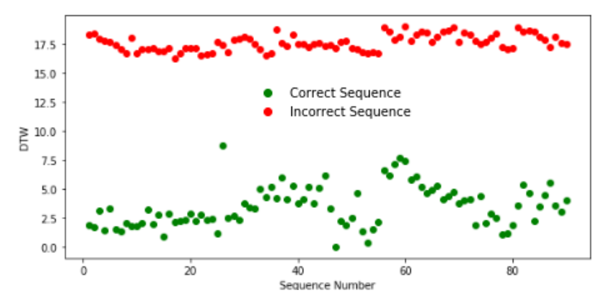

Fig. 2. DTW vs Sequence number - rbf KPCA

As the graph shows, there is a clear split between the incorrect and correct sequences. Hence the separation degree was higher in this case.

B. Scoring Function

The scoring function described in (1) was implemented for the dtw values of the between subjects data for rbf KPCA and is plotted in Figure 3.

A higher score depicts a more correct execution of the exercise. As expected the correct exercise sequences have much higher scores than the incorrect sequences. These scores are fed into the neural network for training as well.

C. Comparison of Neural Networks

The different neural networks that we implemented were a CNN, RNN and a Hybrid model, all of which have been

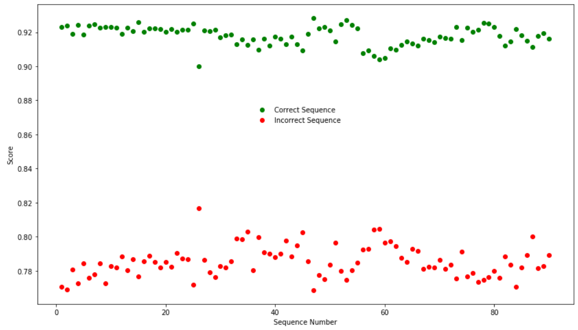

Fig. 3. Score of incorrect and correct exercises

described in Section IV. The results such as the mean absolute deviation and RMS deviation have been displayed in TABLE II.

TABLE II: NEURAL NETWORK RESULTS

Model	Mean Absolute Deviation	RMS Deviation
RNN	0.013	0.027
Hybrid	0.010	0.019
CNN	0.013	0.028

The deviation observed in case of the CNN-LSTM Hybrid model is the least. This implies that the Hybrid model performs the best out of all the other models.

Figure 4 describes the graph obtained by plotting the results of the Hybrid Model for the training dataset. Figure 5 describes

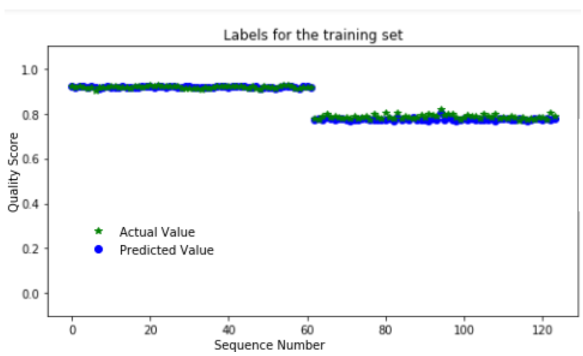

Fig. 4. CNN-LSTM Hybrid model results for training dataset

the graph obtained by plotting the results of the Hybrid model for the test dataset.

The Hybrid model performs the best due to its lower deviation values when compared to the other models.

VI. Conclusion

The experimental results indicate that the movement quality scores generated by the proposed deep learning-based framework closely follow the ground truth quality scores for the movements, and confirm the potential of deep learning models for assessing rehabilitation exercises.Our modification of [1] helped us obtain better deviations. The Hybrid model that we implemented helped us obtain better results. There is scope for

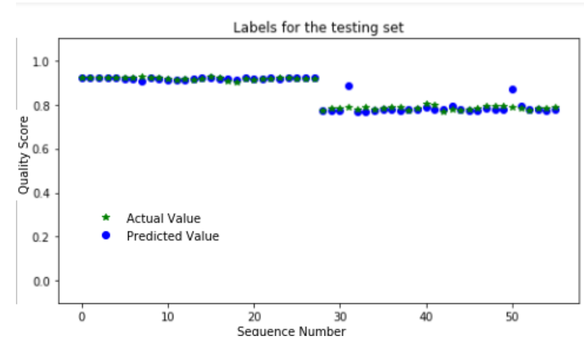

Fig. 5. CNN-LSTM Hybrid model results for test dataset

trying out different and deeper models to increase the obtained accuracy as well.

Our presented research has several limitations. First, the dataset used for validation of the approach comprises rehabilitation exercises collected with healthy subjects, rather than patients in rehabilitation programs. Second, the dataset used is smaller than necessary. There's a necessity to train on a larger dataset before deploying such an application. Third, we lacked the required computation power to process deep networks.

In future work, we will attempt to address the above-listed shortcomings of this study.

References

[1] Yalin Liao and Aleksandar Vakanski and Min Xian,A Deep Learning Framework for Assessing Physical Rehabilitation Exercises, CoRR, abs/1901.10435, 2019

[2] S. R. Machlin, J. Chevan, W. W. Yu, and M. W. Zodet, Determinants of utilization and expenditures for episodes of ambulatory physical therapy among adults, Phys Ther, vol. 91, no. 7, pp. 10181029, Jul. 2011.

[3] K. K. Miller, R. E. Porter, E. DeBaun-Sprague, M. Van Puymbroeck, and A. A. Schmid, Exercise after Stroke: Patient Adherence and Beliefs after Discharge from Rehabilitation, Top Stroke Rehabil, vol. 24, no. 2,pp. 142148, 2017.

[4] P. Maciejasz, J. Eschweiler, K. Gerlach-Hahn, A. Jansen-Troy, and S.Leonhardt, A survey on robotic devices for upper limb rehabilitation,Journal of NeuroEngineering and Rehabilitation, vol. 11, no. 1, p. 3,Jan. 2014.

[5] L. V. Gauthier et al., Video Game Rehabilitation for Outpatient Stroke (VIGoROUS): protocol for a multi-center comparative effectiveness trial of in-home gamified constraint-induced movement therapy for rehabilitation of chronic upper extremity hemiparesis, BMC Neurology, vol. 17, no. 1, p. 109, Jun. 2017.

[6] R. Komatireddy, A. Chokshi, J. Basnett, M. Casale, D. Goble, and T.Shubert, Quality and Quantity of Rehabilitation Exercises Delivered By A 3-D Motion Controlled Camera: A Pilot Study, Int J Phys Med Rehabil, vol. 2, no. 4, Aug. 2014.

[7] D. Antn, A. Goi, A. Illarramendi, J. J. Torres-Unda, and J. Seco, KiReS: A Kinect-based telerehabilitation system, in 2013 IEEE 15th International Conference on e-Health Networking, Applications and Services (Healthcom 2013), 2013, pp. 444448.

[8] A. Vakanski, J. M Ferguson, and S. Lee, Metrics for Performance Evaluation of Patient Exercises during Physical Therapy, International Journal of Physical Medicine Rehabilitation, vol. 05, no. 03, 2017.

[9] H. Sakoe, Dynamic programming algorithm optimization for spoken word recognition, IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 26, pp. 4349, 1978.

[10] A. Vakanski, J. M. Ferguson, and S. Lee, Mathematical Modeling and Evaluation of Human Motions in Physical Therapy Using Mixture Density Neural Networks, J Physiother Phys Rehabil, vol. 1, no. 4, Dec. 2016.

[11] A. Vakanski, H. Jun, D. Paul, and R. Baker, A Data Set of Human Body Movements for Physical Rehabilitation Exercises, Data, vol. 3, no. 1, p. 2, Jan. 2018

[12] Z. Zhang, Q. Fang, L. Wang, and P. Barrett, Template matching based motion classification for unsupervised post-stroke rehabilitation, in International Symposium on Bioelectronics and Bioinformations 2011,2011, pp. 199202.

[13] J.-Y. Jung, J. I. Glasgow, and S. H. Scott, Feature selection and classification for assessment of chronic stroke impairment, in 2008 8th IEEE International Conference on BioInformatics and BioEngineering ,2008, pp. 15.

[14] R. Houmanfar, M. Karg, and D. Kulic, Movement Analysis of Rehabilitation Exercises: Distance Metrics for Measuring Patient Progress,IEEE Systems Journal, vol. 10, no. 3, pp. 10141025, Sep. 2016.

[15] C.-J. Su, C.-Y. Chiang, and J.-Y. Huang,Kinect-enabled home-based rehabilitation system using Dynamic Time Warping and fuzzy logic,Applied Soft Computing, vol. 22, pp. 652666, Sep. 2014

[16] H. Bourlard and Y. Kamp, Auto-association by multilayer perceptrons and singular value decomposition, Biological cybernetics, vol. 59, no.45, pp. 291294, 1988